УДК 004.7:519.2

В.В. Гнатушенко, Н.О. Владимирська
Національна металургійна академія України, Україна
Україна, 49027, м. Дніпропетровськ, пр. Гагаріна, 4

АНАЛІЗ СТАТИСТИЧНИХ ХАРАКТЕРИСТИК КОМУНІКАЦІЙНОЇ ІНФОРМАЦІЇ В КОМП'ЮТЕРНИХ МЕРЕЖАХ

V.V. Hnatushenko, N.O. Vladimirska
National Metallurgical Academy of Ukraine, Ukraine
Ukraine, 49027, c. Dnipropetrovsk, Gagarina av.,4

ANALYSIS OF STATISTICAL CHARACTERISTICS OF THE COMMUNICATION INFORMATION IN THE COMPUTER NETWORKS

В.В. Гнатушенко, Н.А. Владимирская
Национальная металлургическая академия Украины, Украина
Украина, 49027, г. Днепропетровск, пр. Гагарина, 4

АНАЛИЗ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК КОММУНИКАЦИОННОЙ ИНФОРМАЦИИ В КОМПЬЮТЕРНЫХ СЕТЯХ

У статті розглядається задача аналізу статистичних характеристик комунікаційної інформації в комп'ютерних мережах на основі теорії про фрактальність трафіка. Знайдено самоподібну залежність процесів інформаційного потоку на основі експериментальних даних. Для аналізу довгострокової залежності часового ряду використовуються математичні методи з прогнозування фрактальних залежностей з використанням показника Херста.
Ключові слова: статистичні характеристики, трафік, самоподібність, аналіз.

The article considers the problem of analyzing the statistical characteristics of communication information in computer networks based on the theory of fractal traffic. Found self-similar dependence of the processes of information flow on the basis of experimental data. For the analysis of long-term dependence of the time series used mathematical methods to forecast fractal dependencies using Hurst.
Keywords: statistical characteristics, traffic self-similarity analysis.

В статье рассматривается задача анализа статистических характеристик коммуникационной информации в компьютерных сетях на основе теории о фрактальности трафика. Найдено самоподобную зависимость процессов информационного потока на основе экспериментальных данных. Для анализа долгосрочной зависимости временного ряда используются математические методы по прогнозированию фрактальных зависимостей с использованием показателя Херста.
Ключевые слова: статистические характеристики, трафик, самоподобие, анализ.

Основним етапом визначення статистичних характеристик мережевого трафіка є процес вимірювання мережевого трафіка [1-3]. Вимірювання мережі є основою дослідження поведінки мережі, яке включає активне і пасивне вимірювання. На підставі вимірювання дослідник може отримати важливу інформацію про властивості мережевого трафіка. При цьому вимірювання можуть бути проведені різними способами, у різних місцях мережі і в різні періоди часу і тривалості. Місце вимірювання вказує на те, яка частина або елемент КС, а також яка величина вимірюється. При цьому дуже важливо розрізняти вимірювання мережевого трафіка від ідентифікації додатків, оскільки у першому випадку здійснюються збір і обробка даних, а в другому випадку – розпізнавання і класифікація деяких характеристик мережевого трафіка. У свою чергу ідентифікація мережевого трафіка є невід'ємною частиною класифікації, так як класифікація неможлива без його ідентифікації.

Метою даної роботи є дослідження статистичних характеристик комунікаційної інформації мережевих трафіків на наявність різних типів часових залежностей.

Аналіз мережевого трафіка може бути здійснений на кількох абстрактних рівнях: на рівні номерів портів, вмісту пакета, потоку, заголовка пакету і на рівні біта (тобто обсягу трафіка) [2]. При цьому характеристики мережевого трафіка на кожному рівні відрізняються, наприклад, на рівні пакету: мережевий трафік характеризується розміром пакета і часовим інтервалом між пакетами. А аналіз на рівні біта в основному стосується кількісних характеристик мережевого трафіка, таких як інтенсивність передачі і пропускна здатність обміну в каналах. На рівні пакетів розглядається процедура прибуття IP-пакетів, тобто інтенсивність їх затримки і втрати пакетів.

Активний вимір здійснюється шляхом введення стандартних тестових пакетів (наприклад, пакетів ping, traceroute, pathchar, ICMP-протоколів (Internet Control Message Protocol)), щоб виміряти їх поведінку [4]. Однак активний вимір має певні недоліки: по-перше, введення тестових пакетів може впливати на мережевий трафік і порушити нормальне функціонування, що в свою чергу вплине на результат моніторингу; по-друге, одночасний моніторинг великої кількості вузлів, тобто генерація великої кількості тестових пакетів, може порушити нормальне функціонування. На відміну від активного вимірювання, пасивний вимір здійснюється на основі аналізу даних, зібраних з одного або декількох вузлів. При пасивному вимірюванні мережі немає необхідності посилки тестових пакетів, і це означає, що мережевий трафік не порушується. Зазвичай пасивний вимір використовується для моніторингу основних мережевих пристроїв. Недолік пасивного вимірювання мережі пов'язаний з труднощами обробки величезного обсягу даних, зібраних при вимірюванні.

Останні дослідження різних типів мережевого трафіка переконливо показують, що він характеризується властивостями самоподібності і довгостроковими залежностями. Самоподібний трафік має особливу структуру, що зберігається на багатьох масштабах – в реалізації завжди присутня деяка кількість дуже великих викидів при відносно невеликому середньому рівні трафіка. Дане явище значно погіршує характеристики (збільшує втрати, затримки) при проходженні трафіка через вузли мережі. З цього випливає, що методи моделювання і розрахунку мережевих систем, засновані на традиційних припущеннях, які використовуються на даний час, не дають адекватної картини, яка відбувається в мережі.

У багатьох сучасних роботах відзначається, що об'єднання трафіка з декількох джерел призводить до того, що трафік стає сильно автокорельованим з довго-тривалою залежністю [5]. Це призводить до того, що стійкість кореляційних структур не зникає навіть для великих значень. Іншими словами, сукупність безлічі джерел даних, які проявляють синдром нескінченної дисперсії, в результаті дає самоподібний об'єднаний мережевий трафік, який прагне до фрактального броунівського руху. Крім того, дослідження різних джерел трафіка показує, що сильно мінлива поведінка – це властивість, притаманна архітектурі клієнт/сервер.

Не існує єдиного причинного фактора, що викликає самоподібність. Різні кореляції, що існують в самоподібному мережевому трафіку, і які впливають на різні часові масштаби, можуть виникати з багатьох причин, проявляючи себе в характеристиках на конкретних часових масштабах.

Причиною довготривалої залежності в мережевому трафіку можуть бути наступні фактори:

- поведінка користувача і прикладного програмного забезпечення;
- генерація, структура і пошук даних;
- об'єднання трафіка;
- засоби адміністрування мережі;
- механізми оптимізації, засновані на зворотному зв'язку;
- ускладнення структури мережі, збільшення числа абонентів.

У дослідженні проводився моніторинг сервера корпоративної мережі ТОВ «Менора ЛТД». Фізична машина поділена на кілька віртуальних з ОС Linux та ОС Windows, кожна з яких використовується під ряд завдань, таких як СУБД, web-сервери, файл-сервери і т.д. Велика частина трафіка передається по HTTP, інтерфейс 100Мбіт Ethernet.

Структура корпоративної мережі ТОВ «Менора ЛТД» складається з 5 сегментів: локальні мережі «Office», «Hotel», «Restaurant», «Guest» та мережі відеоспостереження.

Кожен фізичний і віртуальний сервер мережі використовує для моніторингу Zabbix. Zabbix – це клієнт-серверний додаток, що використовується для збору, зберігання і обробки інформації про стан мережі, мережеві навантаження, а також стан операційної системи сервера в реальному часі. Додаток широко використовується адміністраторами мережі для моніторингу та своєчасного оповіщення про збої, перевантаження і відмови устаткування. Для подальшої обробки проводилося накопичення статистичних даних наступних параметрів на сегменті мережі «Office» (див. рис.1):

- об'єм кеш-пам'яті;
- об'єм буферної пам'яті;
- час простою процесора при операціях вводу / виводу;
- час процесора в режимі очікування;
- час обробки процесором процесів користувачів;
- час обробки процесором системних процесорів;
- об'єм вільної пам'яті;
- вхідний / вихідний трафік (біт/с);
- сумарне завантаження процесора.

Дані знімалися з різними часовими проміжками для різних параметрів, від 1 до 60 секунд протягом доби. Точку зняття показань ілюстровано на рис. 1.

Часову залежність об'єму трафіка в одиницю часу (біт/с) приведено на рисунку 2 для вхідного і вихідного трафіка. На рисунку 3 показаний ступінь завантаження ЦПУ (у %), а рисунок 4 ілюструє обсяг вільної пам'яті (біт).

Отриманні дані у числовому вигляді були збережені як текстові файли та імпортовані до програми Selfis – програми дослідження властивостей мережевого трафіка у вільному доступі. Оцінка параметра не тільки може допомогти зробити висновок про самоподібність процесу, а й дозволить надалі застосувати до нього ряд математичних методів з прогнозування фрактальних процесів. Виконання короткострокових прогнозів навантаження сервера може допомогти в розробці методів апаратної й програмної оптимізації мережі.

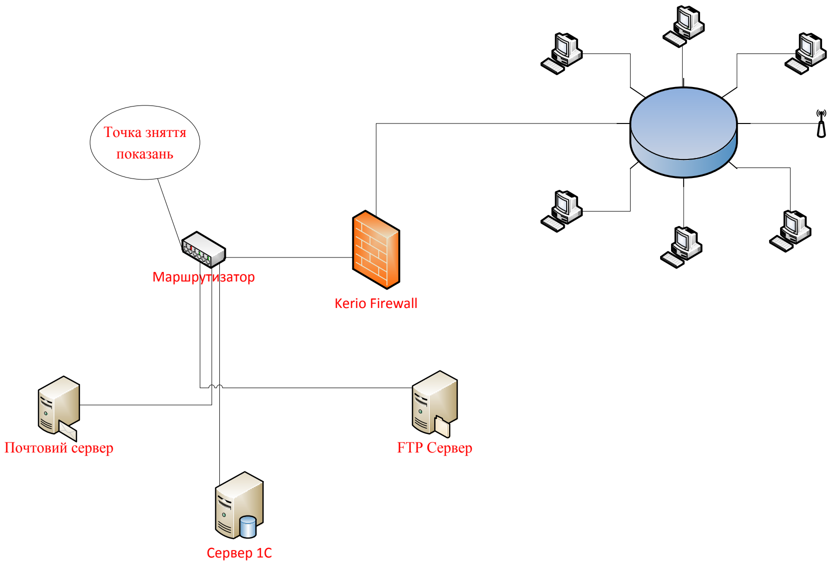
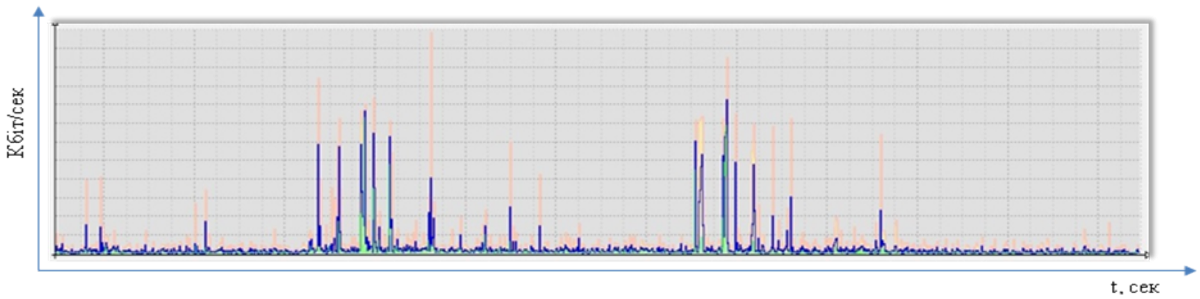
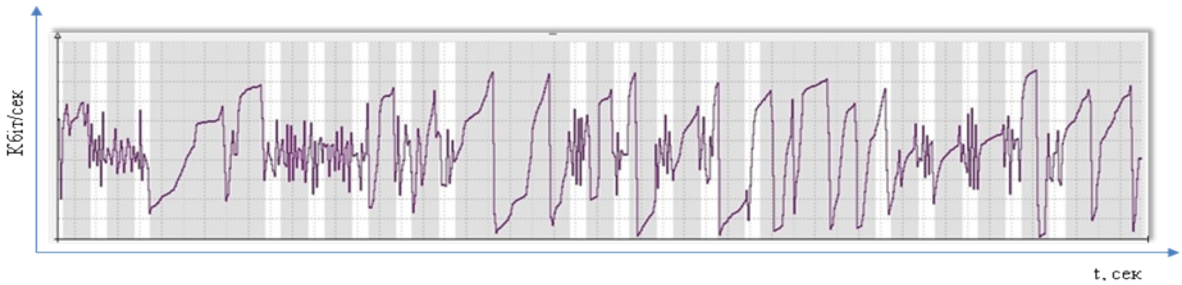


Рис. 1. Сегмент корпоративної мережі, «Office»



а)



б)

Рис. 2. Часова залежність об'єму трафіка в одиницю часу
а) – вхідний трафік, б) – вихідний трафік.

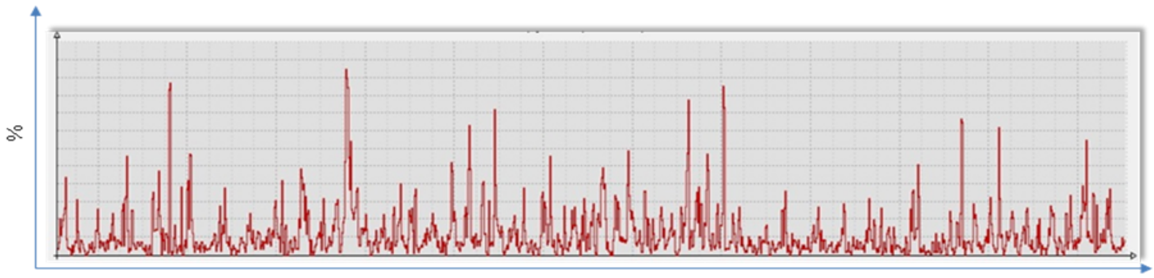


Рис. 3. Ступінь завантаження ЦПУ (%)

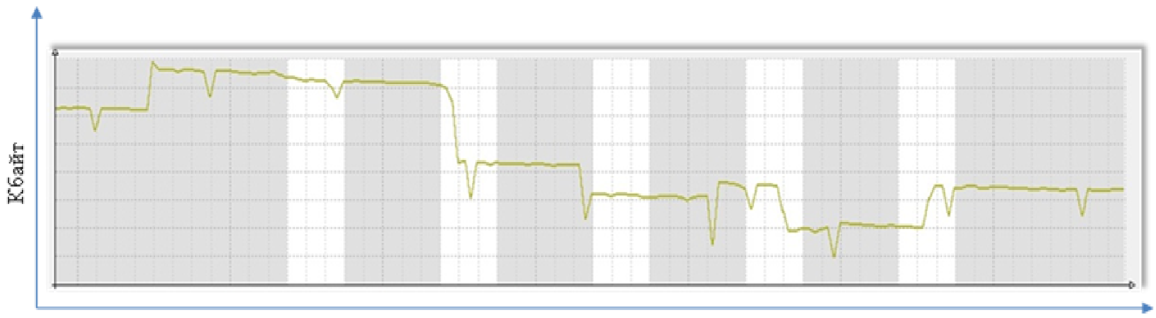


Рис.4. Обсяг вільної пам’яті (біт)

Показник Херста вираховувався різними методами [6]:
- метод сукупної дисперсії (Aggregate Variance);
- метод періодограми (Periodogram);
- метод дисперсії залишків (Variance of residuals);
- R/S метод;
- метод абсолютного моменту (Absolute Moments)

Усі значення параметра Херста для різних характеристик трафіка наведені в таблиці 1.

Таблиця 1. Значення параметра Херста для характеристик трафіка мережі

Характеристика	Параметр Херста				
	Aggregate variance	Periodogram	Variance of residuals	R/S	Absolute Moments
Вхідний трафік (біт/с)	0.576	0.718	0.940	0.555	0.406
Вихідний трафік (біт/с)	0.520	0.586	0.817	0.528	0.176
Завантаження ЦПУ (%)	0.723	0.869	0.910	0.554	1.011
Обсяг вільної пам’яті (біт)	0.708	0.913	1.018	0.350	0.243

Була підтверджена гіпотеза, яка вказує на зв'язок між параметром Херста та інтенсивністю трафіка (таблиця 2). Оскільки у ряді робіт вказується на залежність показника Херста від числа відліків часового ряду, вибірки бралися на рівних інтервалах.

Таблиця 2. Залежність параметра Херста від інтенсивності трафіка

Характеристика	Параметр Херста			
	00:00-06:00	06:00-12:00	12:00-18:00	18:00-24:00
Вхідний трафік (біт/с)	0.576	0.940	0.555	0.406
Вихідний трафік (біт/с)	0.520	0.586	0.817	0.528

За результатами натуральних досліджень найбільшою самоподібністю володіє вхідний трафік у проміжку від 6 до 12 годин. Висока ранкова інтенсивність трафіка пояснюється роботою ряду сервісів, зокрема, документообігу компанії, при якому

автоматична система займається розсилкою документів і наказів в ранкові години. Результати досить повно узгоджуються з результатами відомих числових та експериментальних досліджень.

Дослідження показують, що потоки в сучасних мережах не є простими, мають істотну післядію і самоподібність. Необхідність забезпечення різних категорій мережевих додатків високою якістю обслуговування, обліку затримок, що періодично виникають у передачі даних і втрати пакетів при недостатній продуктивності та обмежених ресурсах пам'яті роблять дослідження статистичних характеристик трафіка дійсно актуальними. У зв'язку з цим важливим питанням є визначення параметрів розподілу інформаційного потоку даних і дослідження властивостей трафіка. Було встановлено самоподібність для всіх процесів, що підтверджує можливість застосування фрактальних моделей для роботи з даними, зокрема для підвищення ефективності роботи сучасних телекомунікаційних систем потрібне створення математичних моделей, що найбільш повно відображають фрактальні властивості мережевих процесів.

Проведені дослідження показали, що зміна параметрів вимірів не тягне за собою якісної зміни показника Херста. Це означає, що для виявлення наявності самоподібності мережевого трафіка не потрібно тривалого збору відомостей про трафік і високої частоти вимірювань. Виконання короткострокових прогнозів навантаження мережі може допомогти в розробці методів апаратної і програмної оптимізації мережі.

Література

1. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов. – СПб.: Питер, 2010. – 672с.
2. Заборовский В.С. Анализ трафика в сетях коммутации пакетов / В.С. Заборовский – СПБб.: СПбГПУ, 2010. – 90с.
3. Нелинейно-динамические модели сетевого трафика / Колесников А. В., Иванов И. П., Басараб М. А. // Нелинейный мир. - 2014. - Т. 12, № 4. - С. 44-56.
4. Буранова М.А. Исследование статистических характеристик самоподобного телекоммуникацион-ного трафика / М.А. Буранова // Инфокоммуникационные технологии. 2012. - Т. 10, № 4. - С. 35-40.
5. Гнатушенко В.В. Данладі Алі. Дослідження самоподібних процесів передачі трафіка на основі ON/OFF моделі/ В.В. Гнатушенко, Алі Данладі// Вісник Національного університету «Львівська політехніка», серія «Комп'ютерні науки та інформаційні технології», – Львів, 2013. – № 751. – С. 87-94.
6. Дж. Скотт Хогдал. Анализ и диагностика компьютерных сетей.– Изд. ЛОРИ, 2001. – 353 с.

Literatura

1. Olifer V.G., Olifer N.A. Computer networks. Principles, technologies, protocols: Trainee for schools. - SPb .: Peter, 2010. – 672 p.
2. Zaborowski V.S. Analysis of traffic in networks packet switching / V.S. Zaborowski - SPb .: Peter, 2010. - 90 p.
3. Nonlinear dynamic model of network traffic / Kolesnikov A.V., Ivanov I.P., Basarab M.A.// Nonlinear world. - 2014. - V. 12, № 4. - pp 44-56.
4. Buranova, M.A. Study of the statistical characteristics of self-similar telecommunication traffic /M.A. Buranova// Information and Communication Technologies, 2012. - V. 10, № 4. - pp 35-40.
5. Gnatushenko V.V., Danladi Ali. Studies of self-similar processes traffic based on the ON/OFF model / V.V. Gnatushenko, Ali Danladi // Bulletin of the National University "Lviv Polytechnic", Series of "Computer Science and Information Technologies" - Lviv, 2013. - № 751. - pp. 87-94.
6. J. Scott Hogdal. Analysis and diagnosis of computer nets.- Ed. Lori, 2001. - 353 p.

РЕЗЮМЕ

В.В. Гнатушенко, Н.А. Владимирская

Анализ статистических характеристик коммуникационной информации в компьютерных сетях

В данной статье рассматривается задача анализа статистических характеристик коммуникационной информации в компьютерных сетях на основе теории о фрактальности трафика. Найдено самоподобную зависимость процессов информационного потока на основе экспериментальных данных. Использование данных математических методов по прогнозированию позволяет по величине показателя Херста определять краткосрочные зависимости временного ряда.

Исследования показывают, что потоки в современных сетях не простые, имеют существенное последействие и самоподобие. Установлено самоподобие для входного и выходного, временной зависимости объема трафика в единицу времени, степени загрузки ЦПУ, объема свободной памяти. Это подтверждает возможность применения фрактальных моделей для работы с данными, в частности для повышения эффективности работы современных коммуникационных систем, а также разкаботки методов аппаратной и программной оптимизации сети.

Показано, что изменение параметров измерений не влечет за собой качественного изменения показателя Херста. Это означает, что для выявления наличия самоподобия сетевого трафика не нужно длительного сбора сведений о трафике и высокой частоты измерений.

V.V. Hnatushenko, N.O. Vladimirska

Analysis of statistical characteristics of the communication information in the computer networks

This article discusses the task of analyzing the statistical characteristics of the communication of information on computer networks based on the theory of fractal traffic. Found self-similar dependence of the processes of information flow on the basis of experimental data. Using these mathematical methods for forecasting enables us to determine the value of the Hurst exponent depending on short-term time series.

Research show that flows in today's networks are not simple, have a significant aftereffect and self-similarity. Self-similarity is established for input and output, the time dependence of the volume of traffic per unit time, the load of the CPU, the memory available. This confirms the possibility of applying fractal models for working with data, in particular to improve the efficiency of modern communication systems, as well as the development of hardware and software techniques to optimize the network.

It is shown that the change in the parameters of the measurement does not entail a qualitative change in the Hurst exponent. This means that, for detecting the presence of the self-similarity of network traffic does not need to collect information long traffic and high frequency measurements.

Надійшла до редакції 30.08.2015

